Safety

Case

Title Safety

Case for Coher

ence-Bound

Autonomous and

Semi-Autonomou

s
Syste
ms

Top-Leve

I Safe

ty Clai

m (C0)

The system

does
not
increa

se
harm
under

uncert
ainty,
stress

, or
human

error.

Clai

m C1

Auth

ority

Never

Escalates

Und

er

Degr

aded

Condit
io

ns
Argu

ment

- Intern

al

cohere
nce is

continu
ously

evaluat
ed

- Auth

ority is
explic

itly
bound
ed

by cohere nce

thresholds

- Thres

hold

breach

results in capa

bility
contra
ctio

n

Evidence

- Deterministic

state machine

- Pre-certified

thresholds

- Hard

ware
interlo
cks

preven
ting
softw

are
overrid
e

Claim

C2 —

Failure

Modes

Are Pred

ictabl
e

and
Audi

table Argu

ment

- No
emer

gent
recove
ry

logic

- All

refus

al
paths
are

explicit

- All

contra
ctions

are
logged

with causal

trace

Evidence

- Refusal taxon

omy

- Audit log

schem
a

- Post-

incident repla

y
capabil
ity

Claim

C3 —

Human

Error

Does

Not Cas

cade
Argu

ment

- Human

commands are

validated again

st

current

system certai

nty

- Unsafe

commands

require
explic

it,
logged

override

- Overr

ide

cannot

bypass
physi

cal
constra
ints

Evidence

- Com

mand
arbitr

ation rules

- Auth

ority
hando
ff

protoc
ol

- Hum

an-in-
the-

loop validation

tests

Clai

m C4

Syst

em
Fails

Con

servat
ive

Iy
Argu

ment

- Perfor

man

ce is
sacrific
ed

before
safet

y

- Under

stress,
the
syste

m
becom
es

less capable,

not
more
aggre

ssive

Evide

nce

- Stress -test

results

- Cons

ervativ
e fallba

ck

modes

- Deterministic

safe states

Safety

Case

Bottom

Line

This

system meets

or

exceeds

the
intent
of

ISO 26262, IEC

61508, and AI

risk frame works

by
making

authority contra

ction
the
prima

ry
safety

mecha
nism.